

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Teacher Training in Measurement and  
Assessment Skills

Buros-Nebraska Series on Measurement and  
Testing

---

1993

## 4. Some Thoughts on Grading Systems and Grading Practices

James S. Terwilliger  
*University of Minnesota*

Follow this and additional works at: <https://digitalcommons.unl.edu/burosteachertraining>

---

Terwilliger, James S., "4. Some Thoughts on Grading Systems and Grading Practices" (1993). *Teacher Training in Measurement and Assessment Skills*. 6.  
<https://digitalcommons.unl.edu/burosteachertraining/6>

This Article is brought to you for free and open access by the Buros-Nebraska Series on Measurement and Testing at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Teacher Training in Measurement and Assessment Skills by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

# Some Thoughts on Grading Systems and Grading Practices

James S. Terwilliger

*University of Minnesota*

## INTRODUCTION

In his role as a discussant of a series of papers on educational evaluation 23 years ago, Scriven (1970) made the following comments:

While the papers this afternoon did not, on the above account, go far enough in the direction of basic evaluation, from another point of view they began at too abstract a level. They contain no discussion at all of the basic method of educational evaluation, one whose use quantitatively swamps any other. I refer to the practice of grading. Like so many other everyday practices, grading has often seemed too humble to merit the attention of high-powered test and measurement people. My feeling is that it is far more important and in more need of help than anything else they work on. Moreover it admirably illustrates the point just made, that the new critics of bad practices are about as irrational as most defenders of the practices. (p. 114)

Unfortunately, little has changed since this observation was made.

## Reference Works on Educational Measurement and Research

A brief review of three standard reference works reveals a general disdain for the topic of grading. The recently published third edition of *Educational Measurement* (Linn, 1989) contains two chapters that

might logically be expected to touch on grading. Chapter 12, entitled "Designing Tests That Are Integrated with Instruction," identifies *attainment decisions* as one of four types of decisions for which tests are employed. The author devotes approximately one-half page (out of 24 in the chapter) to this type of decision and never mentions grading in relation to attainment. Chapter 14, entitled "Certification of Student Competence," provides a lengthy review of statewide competency testing programs and issues associated with standard setting in such programs. The author has nothing to say about the teacher's role in the certification of competence and standard setting as it relates to grading.

Apparently it simply doesn't occur to measurement specialists that classroom teachers are the ones who have the primary responsibility for making attainment decisions and certifying student competence. The terms *grades* and *grading* do not appear in the index of *Educational Measurement* (Linn, 1989).

A second standard reference is the third edition of the *Handbook of Research on Teaching* (Wittrock, 1986). The three chapters in this volume that would logically be linked to grading practices are Chapter 13 ("Teaching Functions"), Chapter 14 ("Classroom Organization and Management"), and Chapter 17 ("Philosophy of Teaching"). None of these chapters contains any reference to grades.

A third somewhat more general reference is the most recent *Encyclopedia of Educational Research* (Mitzel, 1982). In this volume there are approximately 10 pages devoted to the topic Marking Systems. As the title suggests, this summary deals primarily with the purposes of marking and the popularity of various marking systems. The only reference to the process of *assigning grades* is one page that addresses various orientations (criterion referenced, norm referenced, student potential) a teacher may adopt in determining grades. The orientation a teacher adopts is clearly a topic with both philosophic and psychometric importance. (More will be said about this later.) However, the review in the *Encyclopedia* deals primarily with the relative popularity of these orientations as revealed in surveys of teachers.

### Textbooks on Classroom Measurement and Evaluation

A second potential source of information on grading is the textbooks that provide the framework for the education of teachers on matters related to classroom evaluation. Because teachers are almost universally required to assign grades to students and because these grades are commonly defined to reflect the teacher's evaluation of the performance of students on various tests, quizzes, etc. designed by the

teacher, it follows that textbooks on classroom assessment should provide a wealth of practical advice on how to assign grades to students. Alas, such is not the case!

A sample of 12 such texts was examined. This is not a random sample. Rather, it represents all such texts that were easily accessible. It is likely that this set is biased in favor of texts that are most commonly adopted, due to the fact that 5 of the texts have gone through at least three editions.

Table 1 presents a summary that identifies the texts and gives information concerning the length of each and the number of pages on grading. All texts except one (Hills, 1981) contain a single chapter on a variety of issues associated with grading and grading systems. The number of pages in this chapter in relation to the total length of the book is typically quite small, ranging from 4% to 10%. (For the six chapters in Hills, 1981, the figure is 22%.) As shown in the last column of the table, the number of pages devoted to the actual process of assigning grades (as opposed to discussions of various grading and reporting systems) is pitifully small. Only two authors (Hills and Carey) devote more than 10 pages to the actual grading process and half the books devote only 5 or 6 pages to the topic. It seems fair to conclude that, with two possible exceptions, authors of these textbooks on classroom measurement do not attach a great deal of importance to providing teachers with practical advice on grading.

Table 1. Summary of Treatment of Grading in "Standard" Texts on Educational Measurement

	Edition/ Year	Total Pages (Excluding Appendices)	Pages in Grading Chapter	Pages Devoted to Grade Assignment
<sup>a</sup> Ahman & Glock	5th/1975	430	40	6
Carey	1st/1988	415	40	18
Ebel & Frisbie	4th/1986	340	24	8
Gronlund	5th/1985	488	26	9
<sup>b</sup> Hills	2nd/1981	380	84	25
Hopkins & Antes	2nd/1985	465	32	5
Hopkins, Stanley, & Hopkins	7th/1990	470	20	5
Kubiszyn & Borich	2nd/1987	430	18	5
Mehrens & Lehmann	3rd/1984	595	30	5
Nitko	1st/1983	585	24	6
<sup>a,c</sup> Noll, Scannell, & Craig	4th/1979	480	9	5
<sup>c</sup> Popham	2nd/1990	395	12	0

<sup>a</sup> The most recent edition of this text was not available for review.

<sup>b</sup> Hills devotes six chapters to various issues associated with grades and grading. Three chapters deal with the actual grading process.

<sup>c</sup> Grading is covered in a general chapter on the uses of data.



## Research Literature on Measurement

A possible final source of advice on grading is the general literature on educational measurement and/or research. An ERIC search was performed covering the literature for the period from January 1, 1976 through September 30, 1989. A total of 91 references was obtained using the descriptor "Assigning Grades."<sup>1</sup> A careful reading of the abstracts for these 91 references revealed that over half of them (54) did not address or dealt only marginally with assigning grades to students in classroom settings. For example, many of these focus upon issues of evaluating student performance in specific settings (rating systems for college-level writing assignments, using reading journals to improve comprehension of complex texts, etc.) or general student evaluation issues (policies on homework assignments in secondary schools, testing practices of teachers in specific educational settings, etc.). Others deal primarily with curriculum issues, the relationship of grades to student ratings of teachers, etc.

The 37 remaining articles can be classified according to the type of article (empirical study, critique/recommendation) and the educational level (Grades K-12, Postsecondary, Unspecified) to which it is addressed. The results are shown in Table 2. There are two striking features revealed in this table. First, the empirical studies of grading are outnumbered by articles that either critique or recommend grading practices by a 2:1 ratio. Second, half the articles refer to grading at the postsecondary level, and the remaining half are equally split between those that refer to precollege settings and articles that are general with respect to educational level.

The numbers in parentheses in the first column of Table 2 refer to the number of *survey studies*. These studies typically report results based upon responses of teachers in a small group of educational institutions. In each case they employ a self-report instrument designed to determine the popularity of various grading philosophies and practices. Survey results at both the secondary (Terwilliger, 1987) and college level (Prather, Smith, & Kodras, 1981) consistently reveal differences in grading philosophies and practices as a function of the subject matter field.

The differences among disciplines are even more obvious when one examines the articles that focus upon critiques and recommendations related to grading. Seven of the 12 articles at the postsecondary

---

<sup>1</sup>Several other descriptors were employed before selecting this phrase. These resulted in extensive lists of references, most of which have nothing to do with the topic of grading (e.g., using the descriptor "Grades" results in 8,547 references, mostly dealing with research on different grade levels in public schools).

Table 2. Summary of Articles on Assigning Grades

	Type of Article		
	Empirical Study	Critique/ Recommendations for Practice	
Educational Level	Grade Levels K-12	5 (1)	4
	Post Secondary	7 (4)	12
	General/ Unspecified Level	0	9
		12 (5)	25
			37

level and two of the four articles at the K-12 level address grading practices within specific subject matter fields. In each case the author critiques practices or recommends alternative grading strategies that are somewhat unique to instructional methods employed in that field. These range from articles on grading in algebra and engineering courses to courses on personal development and career planning. Two articles (Calhoun & Beattie, 1984; Cohen, 1983) deal specifically with grading practices appropriate for special education students who are in mainstream classes. Advice on how to assign grades in such special circumstances currently is not found in standard texts on classroom measurement and evaluation. The nine articles that are not specific with respect to educational level tend to focus either upon narrow technical issues such as determining boundaries for grading (Aiken, 1983), using computers in assigning grades (Hsiao, 1985), or innovative approaches to grading such as contracts (Klein, 1976).

It would be futile to attempt to synthesize the findings and recommendations offered in the 37 articles in Table 2. The literature on grading is defined more by its diversity than by any universal themes. Differences between educational levels and subject matter fields make generalizations risky, if not meaningless. Yet one gets the sense that the fundamental issues at the heart of grading practices are philosophic, not psychometric, in nature. Perhaps this is why the "high-powered test and measurement people" that Scriven (1970) referred to have so little to say on the subject. Therefore, it may be wise to turn elsewhere for perspectives that can, and often do, influence teachers' grading practices.

## TRADITIONAL GRADING<sup>2</sup> AND PHILOSOPHIC ORIENTATIONS

It is not possible to discuss traditional grading practices in an informed manner without first examining the set of beliefs and assumptions underlying such practices. This is rarely done by advocates of traditional approaches to grading (e.g., authors of textbooks on classroom evaluation). However, philosophic views are discussed at length by a variety of critics of grading, both within and outside the professional educational establishment. Because the views of these critics are not without merit and have a great intuitive appeal to many teachers, they should be examined carefully. Consider the following questions:

1. What purposes do grades serve?
2. What are the costs and benefits of grades?
  - a. To students
  - b. To society
3. On what basis should students be judged?
  - a. What data are relevant?
  - b. How should the data be evaluated?

### Advocates of Traditional Grading

To the question concerning the purposes served by grades, those who support them would likely give two answers. First, grades provide a useful basis for making a variety of important decisions by (and about) individual students. These might include (a) determining promotion and/or graduation, (b) awarding scholarships or special honors, (c) determining eligibility for special programs for the talented, and (d) determining admission into college or other advanced training. Second, grades provide a tangible recognition for excellence in academic pursuits. Such recognition rewards past efforts and encourages future success in learning.

Gardner (1984) has described U.S. education as a sorting-out process:

Americans believe that promise should be recognized at whatever level in society it occurs. They like to think that those future presidents dashing off to school may come from any walk of life.

But as education becomes increasingly effective in pulling the able youngster to the top, it becomes an increasingly rugged sorting-out

---

<sup>2</sup>*Grading* is defined here as the process by which a teacher arrives at a value judgment concerning the quality of a student's achievement of course objectives during a specified period of instruction. Evaluation of performances on single examinations, assignments, projects, etc., are discussed in other papers in this volume.

process for everyone concerned. The schools are the golden avenue of opportunity for able youngsters; but they are also the arena in which less able youngsters discover their limitations. This thought rarely occurred to the generations of Americans who dreamed of universal education. They saw the beauty of a system in which young people could go as far as their ability and ambition would take them, without obstacles of money, social standing, religion, or race. They didn't reflect on the pain involved for those who lacked the necessary ability. Yet pain there is and must be. (p. 79)

With regard to costs and benefits, advocates of traditional grades state that the sorting process which results from grading, although sometimes painful, is ultimately of benefit both to students and to society as a whole. Although grades are admittedly imperfect, they do provide an important basis for a meritocracy. Moynihan (1971) stated this succinctly when he commented:

One of the achievements of democracy, although it seems not much regarded as such today, is the system of grading and sorting individuals so that young persons of talent born to modest or lowly circumstances can be recognized for their worth. (Similarly, it provides a means for young persons of social status to demonstrate that they have inherited brains as well as money, as it were.) I have not the least doubt that this system is crude, that it is often cruel, and that it measures only a limited number of things. Yet it measures valid things, by and large. To do away with such systems of accreditation may seem like an egalitarian act, but in fact it would be just the opposite. We would be back to a world in which social connections and privilege count for much more than any of us, I believe, would like. If what you know doesn't count, in the competitions of life, who you know will determine the outcomes. (p. 4)

It is generally agreed among advocates of grading (at least those who write textbooks on measurement) that the basis for a grade should be the performance or achievement of a student, *not* the effort expended, work habits, character traits, etc. The reason for keeping the basis for grades as "pure" as possible is to minimize the confusion that arises when the meaning of a grade is interpreted. A separate system for recording and reporting teacher judgments concerning student effort, work habits, character traits, etc. is recommended if a school system decides such information is desirable.<sup>3</sup>

---

<sup>3</sup>There is a practical question of how many judgments a teacher should be expected to make and how reliable such judgments are likely to be. This may differ substantially depending upon the setting (e.g., primary grade self-contained classes vs. secondary school classes).

It is further agreed by most advocates of grading that grades should reflect a judgment of achievement with respect to other students (i.e., grades should be norm referenced). This is consistent with the belief that a primary purpose of grades is to differentiate among students as part of an ongoing sorting and decision-making process. This is nicely summarized in the following quote from a colleague who served on a student/faculty committee charged with examining the grading system at the University of Minnesota:

In education, grading represents an information system. Historians perhaps can tell us whether the idea of grades originated from the needs of teachers or the needs of pupils. Current critics can comment on the pro-grading motivation of some administrators and the anti-grading motivation of some students. Such commentary, historical or contemporary, seems not to contribute much to logical analysis. The present social climate encourages a view of academic grading as pejoratively "discriminatory" rather than helpfully "discriminating." The ultimate reality is that Nature does differentiate. Given that fact, we may retreat philosophically from the ensuing pejorative "competition," or we can advance functionally with a helpful "division of labor." (Schofield, 1972.)

Finally, with regard to alternatives (e.g., narrative reports, parent-teacher conferences, contract grading, etc.), advocates of traditional grades consider these to be generally impractical due to time demands that they place on both teachers and those who typically employ grades in decision making. It should be noted, however, that the feasibility of alternatives to traditional grading depends upon the educational context. This will be discussed at greater length in a later section of this paper.

### Critics of Traditional Grading

There are many critics of traditional grading. Three identifiable groups will be discussed. The first comprises individuals who identify strongly with the humanistic movement in education. During the 1960s and the 1970s they advocated fundamental changes in the structure of education and the organization of schools. This movement gave birth to a variety of open or alternative schools in many parts of the United States. A series of publications by Kirschenbaum, Simon, and Napier (1971), Simon and Bellanca (1976), and Bellanca (1977) deal specifically with problems associated with traditional grading and describe alternatives that are thought to be superior to it.

A second source of criticism of traditional grading practices comes from social psychologists and educators who have analyzed

educational practices from the perspective of cooperation versus competition. Deutsch (1979), Johnson and Johnson (1974), and Slavin (1977) have argued that classroom evaluation and reward structures that foster competition among students create an unhealthy environment for learning. They advocate classroom organizations based upon student groups that emphasize teamwork and cooperative learning strategies.

The philosophical premises of this perspective are variants upon humanistic themes. Perhaps the clearest critique of the traditional view of society and grading has been offered by Deutsch (1979):

In addition, I believe we must begin to challenge the assumptions underlying the competitive, meritocratic ideology of our society. We must question whether socioeconomic position in our society is actually distributed on the basis of individual merit. In addition, we must raise issue with the notion that merit belongs solely to an individual, as though its possession were not strongly influenced by social and biological circumstances largely beyond the individual's control. And we must raise doubts about the traditional answer to the question, Who merits merit?—namely, those who have most merit as a consequence of having been more favored with the conditions that foster merit. Finally, we must raise the central question: If the competitive grading system in our schools—a less corrupted version of a competitive merit system than the one that characterizes our larger society—does not foster a social environment that is conducive to individual well-being and effective social cooperation, why would one expect that such values would be fostered in a society that is dominated by a competitive, meritocratic ideology? If the competitive-hierarchical atmosphere is not good for our children, is it good for us? (p. 401)

Research reviews by Johnson and Johnson (1974) and Slavin (1977) conclude that cooperative learning strategies produce achievement outcomes equal to or better than competitive learning approaches in many classroom settings. Further, they conclude that student attitudes toward school and toward peers is much more positive in cooperative learning environments. It should be noted that most of these studies were conducted in elementary schools.

A third group of critics of traditional educational practices has become active in the outcome-based school movement. As reflected in a statement by Spady (1981), this group adopts a strong behavioristic approach to education with an emphasis upon detailed and explicit statements of learning outcomes, mastery-based instructional systems, and criterion-referenced assessment procedures. Spady (1981) lists the following philosophical premises of outcome-based education:

1. Almost all students are capable of achieving excellence in learning the essentials of formal schooling.
2. Success influences self-concept; self-concept influences learning and behavior.
3. The instructional process can be changed to improve learning.
4. Schools can maximize the learning conditions for all students by:
  - a. establishing a school climate which continually affirms the worth and diversity of all students;
  - b. specifying expected learning outcomes;
  - c. expecting that all students perform at high levels of learning;
  - d. ensuring that all students experience opportunities for personal success;
  - e. varying the time for learning according to the needs of each student and the complexity of the task;
  - f. having staff and students both take responsibility for successful learning outcomes;
  - g. determining instructional assignment directly through continuous assessment of student learning; and
  - h. certifying educational progress whenever demonstrated mastery is assessed and validated. (p. 2)

As might be expected, none of the three groups of critics believe that traditional grades serve a useful purpose. Grades are viewed as an artificial and harmful reward system that has little to do with learning. Grades are also seen as a mechanism to exert control over students. Students who learn to please the teacher are rewarded with high grades; students who do not frequently suffer low self-esteem and quit trying. Furthermore, even if grades reflect general learning, they provide little or no information concerning specifically *what* a student has learned.

Critics argue that the costs of traditional grading both to students and to society as a whole far outweigh the benefits. They claim that the disruptive effect of grades upon the educational process cannot be justified by the rather weak relationship of grades to later educational success, although it is admitted that secondary school grades are the best single predictor of college grades. The strength of the typical correlation between secondary and college grades (e.g., .50-.60) is not regarded as having any practical utility. The lack of *any* systematic relationship between grades and indices of success in nonschool settings (i.e., on-the-job performance) is also frequently noted.

Critics vary somewhat with respect to their preference for the proper basis for evaluating students. Humanists are proponents of approaches to evaluation that incorporate as much information as possible about the individual student. For example, they typically recommend that student achievement be judged with respect to the ability or improvement that a student demonstrates. Thorndike (1969a) has referred to this as evaluation with respect to potential. Another approach is "grading by contract." All these approaches individualize the judgment made by teachers and virtually assures all students who made a reasonable effort that they will not fail.

Advocates of cooperative learning strategies are highly critical of norm-referenced assessment and grading, which they regard as the epitome of a competitive system. As an alternative they emphasize group projects in which the assessment of each individual student is heavily dependent upon the quality of the product produced by the student's group. Other factors that determine a student's evaluation might include ratings by peers within the student's group, ratings by peers who are *not* members of the student's group, teacher observations of group interactions, and selected individual achievement data that are independent of group data. The relative weighting of each of these factors varies from one setting to the next. However, the important point is that the grade assigned to each student is influenced by both the performance of the team and the members' perceptions of the contributions made by the student to the team's success.

Advocates of outcome-based education also reject the norm-referenced sorting of students associated with traditional grading. Instead, they propose specific *a priori* statements of learning outcomes against which student performance can be judged. They argue that detailed publicly stated goals provide a more informative basis for evaluation. The criterion-referenced system associated with outcome-based education also is often linked with mastery learning approaches that provide students with multiple trials to demonstrate their competencies. General guidelines for establishing such a system are given by Spady (1981).

Some years ago Ebel (1974) listed 22 arguments (including those cited here) frequently made by the critics of traditional grades. He briefly analyzed each argument and presented a rebuttal. A summary of 8 of the most basic arguments and rejoinders given by Ebel is shown in Figure 1. Readers who wish to pursue this further are encouraged to read Ebel's article in its entirety.



Figure 1. Summary of Eight Criticisms and Rejoinders on Grading

Source: Ebel (1974)

Criticism	Response
1. A single symbol cannot possibly report adequately the complex details of an educational achievement.	Grades aren't intended to provide details. They represent a method of reporting value judgments regarding general level of achievement.
2. The most important outcomes are intangible and hence cannot be assessed or graded.	Important outcomes are, by definition, those that make a difference. With properly constructed measuring devices, differences can be detected and can be the basis for grades.
3. Grades are ineffective motivators of real achievement in education.	Research studies indicate that differential grading does tend to motivate students. It is misleading to imply that high grades and "real achievement" are incompatible. When grades are properly given they are parallel.
4. When students learn mastery, as they should, no differential levels of achievement remain to be graded.	Mastery is difficult to define and does not insure identical levels and types of achievement. In almost any instructional setting some students learn faster and more than do others. This should be reflected in the grade reporting.
5. Low grades may discourage the less able pupils from efforts to learn. Also, some pupils will inevitably fail.	While there can be no guarantee that pupils will not receive low grades, special tutorial and remedial help should be offered to those who receive low grades. No pupil who has taken advantage of such help and made a serious effort to learn should be failed.
6. Grades set universal standards for all pupils despite their great individual differences.	A thoughtful teacher will set standards which are realistic for the class so that the highest grades are achievable. Individual differences in grades are intended to reflect important differences among students.
7. Grading fosters competition rather than cooperation.	Grading emphasizes individual achievement but that does not necessarily imply a competitive learning environment. Many students achieve individual excellence through cooperative learning activities.
8. Grading is more compatible with subject-centered education than with humanistic, child-centered education.	The distinction between subject-centered and child-centered education is not valid. A teacher can recognize his pupils as unique human beings and also help them to achieve subject matter objectives.

## FACTORS THAT INFLUENCE GRADING PRACTICES

## The Temporal Factor

Like all other educational practices, grading practices are influenced by fads and fashions. There are clear cyclical changes in such matters as the choice of the grading system (percent scale, letter grades, pass/fail, etc.) to employ. This is well documented on a general level by Cureton (1971) and in a specific setting by Wrinkle (1947). Little can be learned about the process of grading by studying the popularity of grading systems at any given point in time. The number of categories in grading systems and the symbols that are used may change with time, but these represent somewhat superficial concerns.

On a different level, the influence of various philosophical positions ebbs and flows with the passage of time. The alternative school movement associated with the humanistic view of education became very prominent during the late 1960s and 1970s. Consequently, there was much greater attention during that period to alternative grading practices advocated by humanistic educators. Many schools and colleges modified their grading systems (e.g., replacing "Failure" [F] with "No Credit" [N], providing "Satisfactory/No Credit" [S/N] as an option to letter grades, etc.) and the grade inflation phenomenon was born. For many students, grades were regarded as irrelevant.

More recently, the pendulum has swung back toward a more traditional view. Many of the modifications that were introduced as reforms 20 years ago have been replaced by systems that bear a striking similarity to those that were in place prior to 1960. S/N grading is now less popular and the F has been resurrected in many institutions. In response to grade inflation, a more refined grading system (A+, A, A-, etc.) has been adopted by some colleges in an effort to better differentiate among students. Grades now seem to be regarded as more important by students than they were 20 years ago.

Gardner (1984) has described the situation succinctly in discussing the continuing debate over demands for educational excellence vs. demands for educational equality. Although not identical to issues in grading controversies, there is a substantial overlap in philosophical viewpoints:

If the swings of the pendulum have been excessive at times and the debate more embittered than one might wish, it is because there are extreme and polarizing elements on both sides of the debate.

On the side of quality, the best proponents care deeply about standards and solid subject matter, seek to challenge and stretch the student, and believe that with appropriate adjustments these are suitable goals for students at every level of ability. Unfortunately, also on the side of quality are some who really care only about the college preparatory students and (whether they admit it or not) look down on all the others. Not surprisingly, they give an unpleasant tone to the debate.

On the side of equality, the best proponents care deeply about the economically deprived and about the student of lesser ability—but fully recognize the need for rigorous college preparatory programs. Unfortunately, also on the side of equality are some who are profoundly anti-intellectual, anti-subject matter, and anti-discipline. (p. 89)

### The Educational Level Factor

A critical, but frequently ignored, variable in discussing grading systems and practices is the educational level of the students being evaluated. The importance of educational level follows from the fact that the number and types of decisions made by (and about) students change in significant ways, depending upon the educational and developmental stage at which a student is functioning. The impact of grading upon students is also likely to be different for students at different stages of maturity. For present purposes, four educational levels will be considered: grades K-6, grades 7-12, undergraduate college (13-16), and postgraduate level (e.g., graduate school, medical or law school, other advanced educational programs). Each of these four will be considered briefly.

At the earliest stages (grades K-6) in the educational process, the decisions that are made concerning a pupil's educational progress are very limited. The primary question is, "Has this pupil acquired the basic knowledge and skills typically expected of children at this level?"<sup>4</sup> If the answer is "yes," the decision is to promote the individual to the next level. If the answer is "no," a variety of actions are possible, depending upon the resources available (e.g., do not pro-

---

<sup>4</sup>Naturally, it is assumed that it is reasonable to expect the student to make typical progress. If there is evidence of a serious limitation upon the ability of a child to learn (e.g., certain physical or mental handicapping conditions), it is pointless to hold expectations of typical progress. Under these circumstances, the teacher needs to develop a separate set of expectations that are appropriate to the particular setting. The evaluation of students in such special educational circumstances relies heavily upon judgments of "progress with respect to potential."

mote the pupil and repeat the entire year of instruction, provide intensive remedial instruction during the summer as a condition of promotion, provide special tutorial help concurrent with promotion, etc.).

The limited nature of the options available concerning a pupil's future at this level of the educational ladder argues for a simple system for recording and reporting teacher judgments. At most, it appears that three categories (e.g., Unsatisfactory, Acceptable, Outstanding) are sufficient for communicating to parents. Instead of worrying about more refined distinctions, elementary school teachers could better spend their energy working with individual pupils as they encounter learning problems. Teachers at this level are also in a position (because of self-contained classrooms) to spend a greater fraction of their time monitoring and reporting on the social and emotional development of their pupils. Such matters are clearly a special concern to parents of pupils at this level because problems in the social and emotional domain may have a direct bearing upon learning.

At the next educational level (grades 7-12), the options available to students are typically much more varied than at the earlier stage. The curriculum offers more choices both in terms of subject matter and in terms of special learning opportunities (e.g., accelerated courses, honors programs, work-study opportunities, vocational training, etc.). A student's performance in school during this period plays a major role in determining possible postsecondary job options and/or opportunities for postsecondary education. Entry into higher education is especially significant because this is the gateway to those careers that are generally considered to be the most rewarding, both personally and financially.

The sorting of individuals during the 7th through 12th grades in U.S. education is extremely critical to individual students and to society as a whole. With rare exceptions, the educational choices made during this period of development will, for better or worse, have a profound impact upon opportunities later in life. There is likely to be a continuing debate over whether this is ultimately helpful or harmful to individuals and to society. Nevertheless, there is not a serious debate over whether this is, in fact, the current state of affairs.<sup>5</sup>

---

<sup>5</sup>Critics of traditional grading systems usually fail to recognize that the choices made by (and for) students will be made regardless of whether grades are available. Other sources of information (e.g., standardized test scores) will simply take on more significance as proxy indices of academic talent when grade data are either unavailable or nondiscriminating.

Given the educational system described, a somewhat more refined system for recording and reporting student grades than is used at the earlier levels seems desirable for grades 7-12. For example, a system with five categories (e.g., A, B, C, D, F/N) would provide sufficient differentiation so long as such a system is used properly. That is, there should be a reasonable spread of grades with relatively small frequencies at the extremes and proportionately larger frequencies in the middle category. This does *not* imply that the distribution should be "normal" (or even symmetric) in form. There are bound to be differences from class to class that justify different distribution shapes. However, it would be quite helpful if *written* schoolwide grading policies could be agreed upon that either suggest how grade distributions should look or place general constraints upon what individual teachers can do in assigning grades.

Generally speaking, the issues related to grading at the undergraduate level in college (grades 13-16) parallel those at the secondary level. Students in 4-year undergraduate programs still are faced with a variety of choices with respect to exploring new fields of study, choosing a major field of study, determining whether to pursue advanced study in graduate or professional school, etc. As previously noted, these decisions typically have a long-term impact upon an individual. From the point of view of a meritocratic social system, opportunities offered to students are afforded through a continuation of the sorting that begins at the secondary level. For reasons given above, recommendations concerning the nature of a grading system and how it should be employed in 4-year undergraduate programs are the same as for the secondary level.

There are other postsecondary educational settings where grading systems with fewer categories are appropriate. For example, vocational schools, trade/industrial training programs, and 2-year community college degree programs that are designed to prepare students for specific occupations share a common goal—providing students with the basic knowledge and practical skills necessary to succeed in a specific set of jobs. Here the primary question is, "Does the student possess the knowledge and skills required on the job?" Because the curriculum is ordinarily designed with the specific job demands in mind and students typically are provided with a substantial amount of job-like training as part of the curriculum, competency-based approaches to evaluation are highly appropriate. A grading system comprising no more than three categories will suffice under these conditions (e.g., lacks basic knowledge/skills [unsatisfactory],

possesses basic knowledge/skills [satisfactory], possesses knowledge/skills well beyond the basic level [exceptional]).

Students in graduate and professional school programs are comparable to students in vocationally oriented training in the sense that they are in the terminal stage of their education. Despite the fact that the knowledge base is broader and the cognitive outcomes are more complex at the graduate and professional school level, there seems to be little need for a highly refined grading system. After all, students at this level already have been subjected to extensive sorting and selection prior to their entry into the most advanced stage of their education. Given this fact, the expectation is that almost all students who are admitted will succeed. The major question is, "How well has this student performed in relation to others at the same stage of their education?" No more than three categories for recording judgments should be needed (e.g., unacceptable [U], satisfactory [S], outstanding [O]). Presumably, the first category would be employed rarely, the second category very frequently, and the third category with a fairly low frequency.

#### The Curriculum Factor

A second major variable that should be considered in discussing grades is the role played by a course in the overall curriculum. This is especially important beyond the elementary level because the curriculum becomes more diverse and student choices in selecting course experiences become more varied. For purposes of the present discussion, the curriculum can be partitioned into three major groupings: (a) core academic courses where outcomes are primarily cognitive in nature; (b) specialized courses in disciplines where the outcomes are defined in terms of self-expression in combination with psychomotor and/or affective processes; and (c) general elective courses that emphasize practical skills and/or psychomotor outcomes. The reason for making these distinctions is that performance in courses of different types has different implications for a student's future.

Under the heading of core academic courses at the secondary level are classes in foreign languages, language arts (composition, literature, speech communications, etc.), mathematics (all types), science (biology, chemistry, earth sciences, and physics), social studies (civics, geography, history, etc.), and behavioral/social sciences (psychology, sociology, etc.). All such courses are core in the sense that they present foundational knowledge and concepts that provide a

framework for comprehending the world about us. These courses provide the building blocks for more advanced study at the postsecondary level. Therefore, valid information about how well students perform in these areas is especially crucial to making informed decisions concerning the likelihood of future *academic* success. Historically, grades have been the most valid indicator of future academic performance.

Specialized courses in which outcomes depend largely upon self-expression mixed with psychomotor and/or affective processes are those in the performing arts (dance, drama, musical [instrumental or vocal] performance, etc.), literary arts (writing of fiction and poetry), and visual arts (painting, lithography, sculpture, etc.). Courses of this type are different from core courses in that they rely heavily upon specialized and creative modes of self-expression. More importantly, they are different because they tap aptitudes that have, at best, a marginal relationship to future academic performance as defined by the core curriculum. Valid information about how well students perform in these specialized courses is likely to be useful primarily in predicting future success in the particular field of artistic expression.

There are special problems associated with evaluating artistic performances and creative works. For example, the judgmental standards employed are quite subjective and extremely difficult to define. It is frequently impossible to obtain a clear consensus among experts. To the extent that students are allowed individual discretion in creating performances and projects, there is a fundamental lack of comparability in the finished products. This frequently forces teachers to judge outcomes with respect to individualized expectations based upon beliefs that they hold concerning student talent. Some teachers in artistic fields refuse to make comparative judgments at all because they maintain that each creative work must be judged in terms of how well the artist achieved his/her own creative goals. All of these factors clearly suggest that grading in courses emphasizing artistic expression needs to be treated differently from that in core academic courses.

General elective courses that emphasize practical and/or psychomotor skills include vocational courses (distributive education, home economics, industrial education, etc.) and courses where outcomes relate directly to motor skills (physical education, keyboarding, shorthand, etc.). Obviously, these courses have a different function in the curriculum than do core academic courses. Some of these are designed to provide students with an opportunity to explore special interests and/or to develop practical skills useful in daily life. Others

are designed to give students a preliminary exposure to specific vocational activities. Still others afford an opportunity to engage in active athletic competition. It seems unrealistic to believe that performance in such courses has any predictive relationship to future academic success.

### The Pitfalls of Generalization

The foregoing discussion of the factors that influence grading systems and grading practices underscores the folly of making sweeping recommendations concerning approaches to grading students. Grading methods that are appropriate under one set of circumstances may be highly inappropriate in another setting. Both the number of grading categories employed and the framework used by a teacher in judging performance need to be adapted to the educational context.

Unfortunately, there is a tendency to ignore important situational variables in discussions of grading. The implications of achievement in a core academic course in secondary school for a student's future opportunities are profoundly different than would be the achievement of the same student in home economics or physical education. Likewise, the outcomes of elementary school instruction have very different implications than do the outcomes in a required course for a first-year medical student.

The remainder of this chapter will focus upon grading in the core academic courses at the secondary and college levels. Based upon the premise that grading is an important, albeit distasteful, part of the job of teaching, general principles and specific guidelines for the assignment of grades at the secondary and college levels will be presented.

## THE GRADE ASSIGNMENT PROCESS<sup>6</sup>

### General Principles

There are several general notions concerning grading that should be made explicit at the outset. Some of these ideas are rooted in philosophical beliefs, some come from a cognitive analysis of classroom learning, and others have their origins in classical measurement theory. All are important for teachers to understand if grades are to serve as a defensible basis for decision making.

---

<sup>6</sup>Much of the material in this section is based upon a recent paper by the author (cf. Terwilliger, 1989).



1. Grading is a process of publicly certifying the teacher's judgment of the *quality* of a student's *achievement* in a specific course of study.
2. A teacher's judgment concerning student achievement should be based upon data that have been systematically collected specifically for that purpose. Only data that are directly related to achievement should be employed in grade assignment.
3. Grades should be assigned only as frequently as required by the school or college reporting system. This will allow for the collection of a sufficient amount of data to guarantee that grades are reliably assigned.
4. The assignment of a grade of "Failure" (F) or "No Credit" (N) has special importance. The basis for assigning such a grade should be a *categorical judgment* of the student's performance that is independent of the achievement of other students.
5. Realistic expectations concerning student achievement can only be obtained through experience. Teachers typically arrive at grading practices appropriate to specific settings through a process of trial and error.

The first principle is based upon the assumption that the meaning of a grade is clarified by considering only evidence directly linked to achievement. The utility of grades for decision making is diminished if a teacher attempts to factor in judgments of student effort, potential, work habits, etc. If the reporting system used requires the teacher to make such judgments, these should be recorded and reported separately from the grade.

Further, the quality of achievement in any subject matter should be defined in terms of the *level* of the outcomes achieved by students, not the amount of work students perform.<sup>7</sup> There are several general hierarchical systems for defining the cognitive level of learning outcomes. Perhaps the best known is the *Taxonomy of Educational Objectives* (Bloom, Englehart, Furst, Hill, & Krathwohl, 1956). Another more recent system has been proposed by Presseisen (1986). These are useful for a variety of purposes beyond the assignment of grades.

The second principle assumes that grades are based upon some composite index derived from a clearly defined data base. This means

---

<sup>7</sup>Contract grading schemes defined in terms of quantity of work or the granting of "extra credit" for work beyond that generally required should be discouraged. Such approaches may encourage and reward effort but they have no relationship to evaluation of quality.

that a teacher should have an *a priori* plan for collecting data. The amount of data collected should be sufficient to assure reliability and the variety of data should be sufficient to assure that the basis for judging achievement is broadly defined. No claim is made that this results in objective grades. However, it does make the grading process more explicit.

Although grades should be clearly linked to data, it does not follow that all data collected by a teacher need to be considered when assigning grades. There are other reasons for collecting data (e.g., giving periodic feedback to students, providing practice exercises, problems, quizzes, etc.) and obtaining data for the purpose of evaluating instruction, course materials, etc.

The third principle assumes a fundamental distinction between the process of judging performance and the process of data collection. It is well known that the validity of a judgment is enhanced if the data employed are reliable and relevant. Both reliability and relevance are improved when a substantial amount of data are collected over an extended period of time.

Critics of traditional grading are correct in saying that the importance of grades in the minds of students is frequently exaggerated. This is due, in part, to the inappropriate use of grades. The teacher who falls into the trap of assigning grades every time class assignments are due, quizzes are administered, projects are completed, tests are given, etc. is only contributing to many of the negative side effects of grades noted by Kirschenbaum et al. (1971) in *Wad-ja-get?* Teachers should learn to differentiate clearly between the act of making a judgment (assigning a grade) and the act of collecting data (obtaining information on which to base a judgment). Data collection should occur with much greater frequency than grading.

The fourth principle addresses the painful issue of failure or no credit. This is usually the "worst case" scenario for both a student and a teacher. The only way to avoid such a scenario is to refuse to consider a grade of F or N as an option. Some critics of grading endorse that approach. Whatever short-term benefit this has for the student may result in a long-term cost both to the student and to society (e.g., the student may subsequently be in a more advanced course or a job setting where unlearned knowledge and/or skills are critical).

The method for determining a grade of F or N should be as fair and honest as possible. *Fair* means that students know exactly what performance expectations define the boundary between F and "non-F." *Honest* means that the performance expectations are established

by the teacher based upon a thoughtful and thorough specification of the knowledge and/or skills that are regarded as minimal or essential outcomes of the course. A grade of F should result from the teacher's judgment that a student does not possess a minimal level of competence as defined by the essential course outcomes. In other words, failing grades should be assigned on the basis of a categorical (criterion-referenced) judgment rather than a comparative (norm-referenced) judgment.

The final principle is an acknowledgement that grading practices evolve with experience. Ideas regarding what data to collect, how to design assignments, tests, etc., and how to use the results have to be developed. Performance standards are established and modified in an iterative fashion. Norms, whether they be explicit or in the teacher's head, are built from long experience with different groups of students. In summary, developing a practical and valid set of grading practices is a long-term undertaking.

### A Specific Approach to Grading

The five general principles discussed above provide a general framework for thinking about grade assignment but they do not provide specific guidance. This section will describe in detail an approach to grading that can be adapted to a wide variety of classroom settings. Prior to doing so, however, there are two specific recommendations that will improve grading practices regardless of the particular approach employed:

1. At the beginning of each term a teacher should prepare an evaluation outline for distribution to students. This outline should give dates for quizzes, exams, class presentations, etc. as well as due dates for assignments and projects. In addition, the outline should specify the nature of the quizzes and examinations (choice response vs. free response questions) and conditions under which they are to be administered (time limits, use of reference materials, etc.). Finally, the outline should clearly indicate the relative weight to be given to each item of data in arriving at grades.
2. All data to be employed in grading should be expressed in *quantitative* form. This implies that a teacher designs a scoring system, however primitive, for counting points earned on all quizzes, exams, assignments, projects, presentations, etc. The teacher should provide feedback to students in terms of

points earned rather than letter grades or some corresponding evaluation of performance.

Both of these recommendations are based upon the assumption that the teacher has acquired substantial experience with the subject matter in question. Therefore, these should be viewed as “end state” conditions after the teacher has experimented with different methods for collecting, coding, and aggregating data relevant to achievement in the subject matter.

*Minimal vs. developmental objectives.* In every subject matter and educational level, there are instructional outcomes that are essential or basic in the sense that they define the most rudimentary knowledge and skills. In principle, these are outcomes that every student is expected to achieve. Gronlund (1985) refers to such outcomes as *minimal objectives*.

In contrast to minimal objectives, in any subject matter and educational level there are also a large (and unspecified) number of instructional outcomes that define more complex and advanced levels of achievement. In principle, these more advanced outcomes are attainable only after students have mastered the minimal objectives. However, due to their diverse and subtle nature, it is *not* assumed that all students will achieve all (or even most) of them. Consequently, it is expected that there will be reliable individual differences among students with respect to performance on these more advanced outcomes. Gronlund (1985) has called these *developmental objectives* because they reflect a student’s level of development in striving to achieve the more challenging instructional outcomes.

The distinction between minimal and developmental objectives is crucial not only to the assignment of grades but also to designing instructional systems. For example, Gronlund (1973) argues that Bloom’s (1968) notions about mastery learning and mastery testing apply well to minimal objectives but are not as appropriate in the case of developmental objectives. The same distinction holds for all approaches to instruction (e.g., outcome-based or competency-based education) that emphasize all students achieving at the same a priori standard.

There are several ways to differentiate minimal from developmental objectives. For example, Gronlund (1973) defines minimal objectives in terms of the following questions:

1. What minimum knowledge and skills are prerequisite to further learning in the same area (e.g., knowledge of terms, measurement skills)?

2. What basic skills are prerequisite to learning in other areas (e.g., reading skills, computational skills, language skills)?
3. What minimum skill is needed for safe performance in some particular activity (e.g., using laboratory equipment, driving an automobile)?
4. What knowledge and skills are needed to attain minimum job proficiency (e.g., lathe operation, typing skill)?
5. What minimum knowledge and skills are needed to function in everyday, out-of-school, situations (e.g., reading, writing, speaking)? (p. 8)

Gronlund (1973) further suggests that the definition of outcomes that all students are expected to master be done cooperatively by teachers in consultation with subject matter authorities, curriculum specialists, and experts on learning.

A second way to distinguish minimal from developmental outcomes is to refer to cognitive analyses of instruction. Minimal objectives correspond to lower level cognitive outcomes, whereas developmental objectives correspond to higher level cognitive outcomes. For example, Presseisen (1986) describes four categories of *thinking skills*:

- a. Essential cognitive processes—the basic thinking skills that are the building blocks of thought development;
- b. Higher-order cognitive processes—the more complex thinking skills, which may be harder to define but which are based on the essential cognitive processes;
- c. Metacognitive processes—the learning to learn skills aimed at making thinking more conscious and the student more aware of the ways one can go about problem solving or decision making; and
- d. Epistemic cognitive processes—the kinds of thinking related to particular bodies of knowledge or subject matters and the particular problems addressed by these knowledge areas as well as the interdisciplinary relationships among content areas. (p. 9)

The first category might serve as a basis for defining minimal objectives, whereas some mixture of metacognitive and epistemic process could define developmental objectives. This is supported by Presseisen's (1986) description of the difference between the first two categories:

There is a decided difference between what is meant as a higher-order thinking and the exact, standardized, minimal competency objectives often included in basic skills instruction. Simplistic, rote information that fits limited instructional sequences is not sufficient

as the material upon which to develop students' higher-order thinking. (p. 11)

Terwilliger (1989) has argued that *novelty* is a useful basis for distinguishing minimal from developmental objectives:

One concept that I employ is novelty. I believe that outcomes that are defined as minimal objectives are those that test students' ability to deal with familiar concepts and rehearsed skills. By definition, such outcomes have a low level of novelty. In contrast, developmental objectives test students' ability to apply learning to new material or situations. (p. 17)

It can also be noted that the application of learning to new settings has historically been described as *transfer* of learning. Many years ago Thorndike (1969b) described this as the basis for teaching and testing for understanding:

The crucial indicator of a student's understanding of a concept, a principle, or a procedure is that he is able to apply it in circumstances that are different from those under which it was taught. Transferability is the key feature of meaningful learning. So if we are to test for understanding, we must test in circumstances that are at least in part new.

Does a child really know how to read a map? Try him with one that is different from the one in the book. Does he really understand denominate numbers? Give him some problems phrased in "wugs," "pogs," and "pilzits," the units used in measurement in the country of "Zoolumbia." (I hope that a real "Zoolumbia" hasn't sprung into existence recently without my being aware of it.) Does the Bill of Rights mean anything to him except a lot of words to be memorized? Ask him in what way recently proposed laws to regulate the sale of firearms might be considered unconstitutional. (p. 2)

*Minimal outcomes and failure.* A series of special quizzes, exercises, etc. should be designed to measure student achievement of the minimal objectives. These assessments function like mastery tests in the following ways:

1. Some a priori performance standard (for instance, 75% or 80% of maximum possible) is set for each assessment.
2. The expectation is that most, if not all, students (for instance, 90-95%) will perform at or above the level specified by the standard.
3. Students who fail to achieve at or above the standard will be given a second opportunity to take a parallel version of the quiz, exercise, etc. after review and remediation. The *higher* of

the two scores achieved (original vs. parallel version) will be recorded for the student.

It is important to compare the *long-term* failure rate on each minimal objectives assessment with the expectation that 90-95% of the students will achieve the minimal objectives. The statement of such an expectation provides a benchmark for determining if the difficulty level of the minimal objectives assessment is appropriate. Failure rates may be quite high for some quizzes, etc., suggesting that either they are too difficult or the standard is too high. For other measures failure rates may be zero, suggesting that either the learning outcomes are somewhat trivial or the standard is too low. It is important that the difficulty level of measures of minimal objectives outcomes be properly calibrated with the standard set for pass/fail decisions. This usually requires two or more administrations of a measure.

Warren (1971) has made the following insightful comment with regard to the setting of "absolute" standards:

Even in the British system of external examiners and in criterion-referenced testing, the "absolute" standard is established in relation to some expectation of performance based on past experience with examinees in similar circumstances. The real issue is in specifying the source of the standard on which grades are to be based. (p. 23)

An aggregate score on all minimal objectives assessments is determined for each student at the time grades are to be reported. The score typically will be expressed as a percent of the maximum possible points on all minimal objectives assessments administered during the grading period. Pass/fail decisions should be made by comparing the aggregate score of each student to the a priori standard. Those who achieve the standard "pass" and those who do not "fail."

*Developmental objectives and passing grades.* A separate set of achievement measures must be developed as a basis for differentiating levels of acceptable performance. These measures define differences among students in their achievement on the cognitively more complex developmental objectives. No a priori standard is specified for these measures. Instead, the performance of each student is interpreted with respect to norms derived from the administration of developmental objectives measures to reasonably large groups (for instance, 50 or more) of students. Normative data can be built up over time where class sizes are small.

As previously noted, measures of developmental objectives should require students to apply knowledge and skills to novel settings. According to Fleming and Chambers (1983), this is *not* what teachers are accustomed to doing. Context-dependent questions that incorporate graphs, diagrams, tables, maps, etc. are useful devices for measuring cognitively complex outcomes. Teachers clearly need much more training than they currently receive in developing questions of this type. Teachers also need more practice in designing assignments, projects, term papers, etc. that require students to engage in critical analyses of novel situations, to integrate and synthesize familiar information with new data, to judge the merits of competing interpretations and contradictory evidence, etc. Activities such as these impress upon students the difference between low-level and high-level outcomes.

It is assumed that properly designed measures of higher order outcomes will result in score distributions in which the average score with respect to the maximum possible is much lower than for measures of minimal objectives. Also, the distribution of scores should be much more symmetrical in form with substantial variability. A summary of the expected statistical properties of the two types of measures is given in Table 3.

Table 3. Expected Characteristics of Score Distributions Resulting From Minimal Objectives and Developmental Objectives Measures

Distribution Characteristic	Minimal Objectives	Developmental Objectives
Shape	Definite negative skew	Approximately unimodal symmetric
Central tendency (difficulty level)	Mean score well above a priori standard (e.g., .05 to .10) when divided by maximum possible score	Mean score divided by maximum possible score in interval between .50 and .70
Variability	Can be small or large; depends primarily upon degree of skew in distribution	Should be quite large



Grades should be assigned on the basis of composite scores that combine data on several developmental objectives measures.<sup>8</sup> (Presumably, the relative weight associated with each measure has been specified for students in the evaluation outline previously mentioned.) Assuming the composite score distribution is as expected, a norm-referenced basis for assigning grades can be employed readily. This is bound to be a trial-and-error process much like that for deciding standards for minimal objectives measures. However, with experience, teachers can develop very explicit norms that can be shared with students to help them understand the basis for judgments being made.

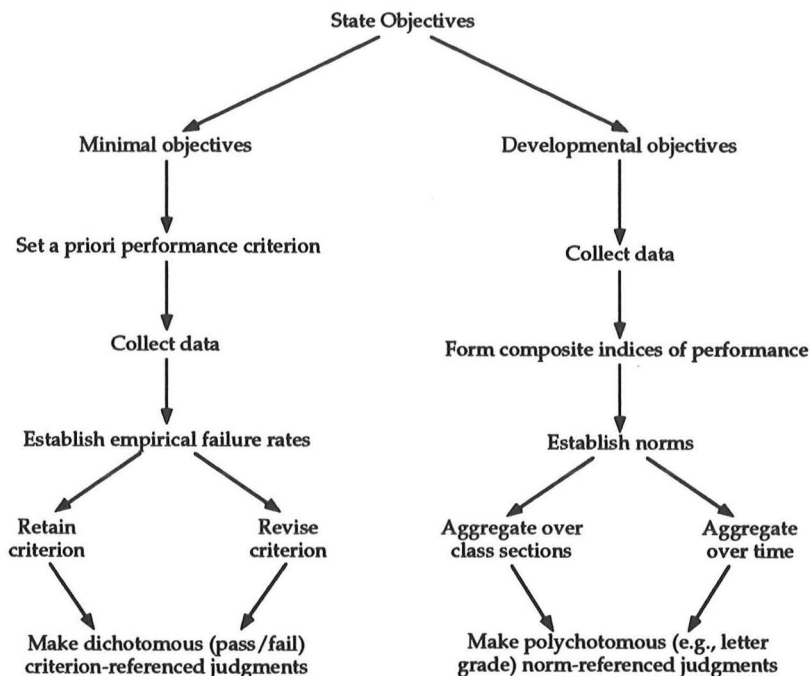
Two comments concerning norm-referenced grade assignment are in order. First, norm referencing does *not* imply a normal curve model. It is helpful if the distribution of composite scores is approximately unimodal and symmetric, but the main concern is that the variation is sufficiently great to assure reliable differences as the basis for grade assignment. Second, critics of grading often equate norm-referenced systems with direct competition among students. This is only the case when the norm group is restricted to others in the same class. The recommendation here is that the norms be based upon a more inclusive group (e.g., all students who have enrolled in the course over a specified period of time, for instance, during the most recent 3-5 years). This will result in much more stable norms and greatly reduce the competitive aspect of grades.

A summary of the grading process that has been described is presented in Figure 2. This makes it very clear that a two-track approach to evaluation is being proposed. One track leads to a dichotomous (pass/fail) decision employing a criterion-referenced model. The second track leads to a polychotomous (e.g., letter grade) decision employing a norm-referenced model. In courses where only pass/fail grades are used, the criterion-referenced model will suffice. In courses where students have the option of enrolling either on the pass/fail or traditional grading system, those on the pass/fail system are required to demonstrate achievement only at the minimal objectives level. Those enrolled on the traditional system must complete all assessment measures. For those students, grading is a two-stage process. First, students must demonstrate mastery of minimal objectives. Then, based upon performance on measures of developmental objectives, grades are assigned using norms.

---

<sup>8</sup>Technical issues associated with weighting measures in the formation of composites are not discussed here. Terwilliger (1977) and Oosterhof (1987) provide detailed treatments of this topic.

Figure 2. Overview of Classroom Evaluation and Grade Assignment



CONCLUSION

Anyone who carefully examines the literature on grading systems and practices is struck by the continuing controversy over grades. Warren (1975) noted that recurring arguments over the purposes and definition of grades can be traced back to the period shortly after the turn of the century. Philosophical differences are at the heart of the controversy.

The scant attention given to the topic by authors of texts on classroom measurement tends to focus on practical and psychometric concerns. The recommendations given in these texts presuppose that teachers accept traditional grading as beneficial both to individual students and to society. However, there is substantial evidence that this is not the case.

Stiggins, Frisbie, and Griswold (1989) report that the majority of secondary teachers they studied employ grading practices that are at variance with the conventional wisdom offered in textbooks on mea-

surement. In discussing the research implications of their findings, Stiggins et al. (1989) refer to philosophical beliefs. They state:

It is a matter of educational values, for example, what information the grade assigned to a student should convey: achievement relative to others (norm referenced) or achievement relative to some absolute performance standards (criterion referenced). No research studies can help to answer the question, Which meaning should grades convey? A teacher's judgment about the grading approach to be used should be dictated by the broader educational values (particularly the theory of teaching) that he or she holds. Until the teacher decides what meaning the grades should convey, most other decisions about grades and grading practices cannot be made. The significant research questions that need to be examined differ between these two grading approaches and even between methods within each approach. (p. 11)

The two-track approach to grade assignment that is recommended here attempts to demonstrate that teachers do not have to choose between criterion-referenced and norm-referenced approaches. Both can (and usually should) be employed in assigning grades. Each approach is uniquely suited to a particular problem faced in assigning grades.

At one level it is possible to obtain an empirical answer to the question concerning the merits of norm-referenced versus criterion-referenced grades. One simply has to compare the predictive power of grades assigned by the two approaches using criteria defined by subsequent performance in academic and/or employment settings. Of course, at a more fundamental level, the question of the relative costs and benefits of grades to students and society cannot be resolved by empirical research, no matter how sophisticated the methodology. Therefore, despite the virtues of any specific set of recommendations, it is safe to assume that the controversy over grades will continue.

## REFERENCES

- Aiken, L. R. (1983). Determining grade boundaries on classroom tests. *Educational and Psychological Measurement*, 43(3), 759-762.
- Bellanca, J. A. (1977). *Grading*. Washington, DC: National Education Association.
- Bloom, B. S. (1968). Learning for mastery. *Evaluation Comment*, 1(Whole 2).
- Bloom, B. S., Englehart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: Handbook I: Cognitive domain*. New York: D. McKay.

- Calhoun, M. L., & Beattie, J. (1984). Assigning grades in the high school mainstream: Perceptions of teachers and students. *Diagnostic*, 9(4), 218-225.
- Cohen, S. B. (1983). Assigning report card grades to the mainstreamed child. *Teaching Exceptional Children*, 15(2), 86-89.
- Cureton, E. L. (1971). The history of grading practices. *National Council on Measurement in Education: Measurement in Education*, 2(Whole 4).
- Deutsch, M. (1979). Education and distributive justice: Some reflections on grading systems. *American Psychologist*, 34(5), 391-401.
- Ebel, R. L. (1974). Shall we get rid of grades? *National Council on Measurement in Education: Measurement in Education*, 5(Whole 4).
- Fleming, M., & Chambers, B. (1983). Teacher-made tests: Windows on the classroom. In M. E. Hathaway (Ed.), *Testing in the schools*. San Francisco: Jossey-Bass.
- Gardner, J. (1984). *Excellence: Can we be equal and excellent too?* New York: Norton.
- Gronlund, N. E. (1973). *Preparing criterion-referenced tests for classroom instruction*. New York: Macmillan.
- Gronlund, N. E. (1985). *Stating objectives for classroom instruction* (3rd ed.). New York: Macmillan.
- Hills, J. R. (1981). *Measurement and evaluation in the classroom* (2nd ed.). Columbus, OH: Merrill.
- Hsiao, F. S. (1985). The electronic gradebook: An application of dBASE II program for academia. *Collegiate Microcomputer*, 3(1), 59-67.
- Johnson, D., & Johnson, R. (1974). Instructional goal structure: Cooperative, competitive, or individualistic. *Review of Educational Research*, 44, 213-240.
- Kirschenbaum, H., Simon, S., & Napier, R. (1971). *Wad-ja-get? The grading game in American education*. New York: Hart.
- Klein, J. T. (1976). Sign now, pay later: Further experiments in student grading. *Exercise-Exchange*, 21(1), 9-12.
- Linn, R. L. (Ed.). (1989). *Educational measurement* (3rd ed.). New York: American Council on Education & Macmillan.
- Mitzel, H. E. (Ed.). (1982). *Encyclopedia of educational research* (5th ed.). New York: Free Press.
- Moynihan, P. (1971). Seek parity on educational achievement, Moynihan urges. *Report on Educational Research*, 3, 4.
- Oosterhof, A. C. (1987). Obtaining intended weights when combining students' scores. *Educational Measurement: Issues and Practice*, 6(4), 29-37.

- Prather, J. E., Smith, G., & Kodras, J. E. (1981). *Implications of differential faculty grading standards and practices by academic field*. Paper presented at annual forum of Association for Institutional Research, Minneapolis, MN. (ERIC Document Reproduction Service No. ED 205 074)
- Presseisen, B. Z. (1986). *Thinking skills: Research and practice*. Washington, DC: National Education Association.
- Schofield, W. (1972). Memo to members of subcommittee on grading, University of Minnesota Assembly Committee on Educational Policy.
- Scriven, M. (1970). *Proceedings of 1969 Invitational Conference on Testing Problems* (pp. 112-115). Princeton, NJ: Educational Testing Service.
- Simon, S. B., & Bellanca, J. A. (Eds.). *Degrading the grading myths: Primer of alternatives to grades and marks*. Washington, DC: Association for Supervision and Curriculum Development.
- Slavin, R. (1977). Classroom reward structure: An analytical and practical review. *Review of Educational Research*, 47, 635-650.
- Spady, W. G. (1981). *Outcome-based instructional systems* (document distributed by the Network for Outcome-Based Schools). Arlington, VA: American Association of School Administrators.
- Stiggins, R. J., Frisbie, D. A., & Griswold, P. A. (1989). Inside high school grading practices: Building a research agenda. *Educational Measurement: Issues and Practice*, 8(2), 5-14.
- Terwilliger, J. (1987). *Classroom evaluation practices of secondary teachers in England and Minnesota*. Paper presented at meeting of National Council on Measurement in Education, Washington, DC.
- Terwilliger, J. S. (1977). Assigning grades: Philosophical issues and practical recommendations. *Journal of Research and Development in Education*, 10(3), 21-39.
- Terwilliger, J. S. (1989). Classroom standard setting and grading practices. *Educational Measurement: Issues and Practice*, 8(2), 15-19.
- Thorndike, R. L. (1969a). Marks and marking systems. In *Encyclopedia of Educational Research* (4th ed., pp. 759-766). New York: The Macmillan Co.
- Thorndike, R. (1969b). Helping teachers use tests. *Measurement in Education*, 1(Whole 1).
- Warren, J. R. (1971). *College grading practices: An overview* (Report No. 9). Washington, DC: ERIC Clearinghouse on Higher Education. (ERIC Document Reproduction Service No. ED 047 164)
- Warren, J. R. (1975). *The continuing controversy over grades* (TM Report 51). Princeton, NJ: ERIC Clearinghouse on Tests, Measurement,

and Evaluation. (ERIC Document Reproduction Service No. ED 117 193)

Wittrock, M. C. (Ed.). (1986). *Handbook of research on teaching* (3rd ed.). New York: Macmillan.

Wrinkle, W. L. (1947). *Improving marking and reporting practices in elementary and secondary schools*. New York: Holt, Rinehart & Winston.

